Runhui Li

Briefing

5+ years' experience in developing distributed systems and hacking source codes of open-sourced distributed systems (e.g., Hadoop). Familiar with C, C++, Python and Java programming in Linux/Unix environments.

Working Experience

2016–Now **Researcher**, *Huawei Future Network Theory Laboratory*.

2015–2016 **Postdoctoral Fellow**, *Department of Computer Science and Engineering*, Chinese University of Hong Kong.

Interests: Big data storage and processing, In-memory storage

Education

2011–2015 **PhD**, *Computer Science and Engineering*, Chinese University of Hong Kong.

Thesis Title: Enabling Efficient and Dependable Clustered File Systems through New Erasure Coding Techniques

2007–2011 **Bachelor of Engineering**, *Computer Science and Technology*, University of Science and Technology of China.

Projects and Open Source Softwares

2016 **Repair Pipelining for Erasure-Coded Storage**
My contribution: Full design and implementation
Briefing: Repair pipelining is a middleware sit on top of distributed file systems (e.g., HDFS, QFS). By fully utilizing network resources, it reduces repair time of erasure-coded storage system by over **90%**.
Project website: will appear after paper getting accepted.

2016 **ADN (Application-Driven Network) Prototype**
My contribution: Lead design and implementation of control plane
Briefing: ADN provides improved QoS guarantee by serving differentiated applications with dynamically-allocated network resources as well as dedicated transmission protocols.

2015 **Recovery-Oriented STAIR Codes in Storage Clusters**
My contribution: Lead implementation

Unit 335-337, Core Bldg – Hong Kong Science Park – Hong Kong
☎ *(+852) 9517 4226* • ✉ *lrhdiy@gmail.com* • *rhli.github.io*

Briefing: R-STAIR code is a storage-efficient code achieves high recovery performance for single failures as well as tolerance for burst failures. R-STAIR code outperform the state-of-the art coding schemes (proposed by Facebook and Microsoft) by **3.6**× in the single failure recovery performance.
Project website: `http://ansrlab.cse.cuhk.edu.hk/software/rstair/`
GitHub repository: `https://github.com/rhli/hadoop-rstair/`

2015 **EAR: Encoding-Aware Replication in Clustered File Systems**
My contribution: Full design and implementation
Briefing: EAR enables efficient and reliable transition from replication to erasure coding. EAR boosts encoding performance by up to **120%** compared with conventional random replication.
Project website: `http://ansrlab.cse.cuhk.edu.hk/software/ear/`
GitHub repository: `https://github.com/rhli/hadoop-EAR/`

2014 **Degraded-First Task Scheduler for MapReduce in Erasure-Coded Storage Clusters**
My contribution: Full design and implementation
Briefing: Degraded-First scheduler optimizes MapReduce performance in presence of component failures. It reduces the runtime of MapReduce job by up to **48%**.
Project website: `http://ansrlab.cse.cuhk.edu.hk/software/dfs/`
GitHub repository: `https://github.com/rhli/Degraded-First-Scheduler/`

2013 **CORE: Regenerating-Coding-Based Recovery for Single and Concurrent Failures**
My contribution: Lead full design and implementation
Briefing: CORE optimizes multi-node failure recovery of regenerating codes. The recovery throughput can be as high as **2.33**× that of widely deployed RS codes.
Project website: `http://ansrlab.cse.cuhk.edu.hk/software/core/`

2011 **NCFS: Network-Coding-Based Distributed File System**
My contribution: Implementation in file system layer
Briefing: NCFS is a proof-of-concept prototype of a Network-Coding-based Distributed File System. NCFS is a proxy-based file system that interconnects multiple storage nodes.
Project website: `http://ansrlab.cse.cuhk.edu.hk/software/ncfs/`

Publications

2017 Runhui Li, Xiaolu Li, Patrick P. C. Lee and Qun Huang
Repair Pipelining for Erasure-Coded Storage.
The 2017 USENIX Annual Technical Conference (ATC'17), Santa Clara, California, July 2017

2017 Eman Ramadan, Arvind Narayanan, Zhi-Li Zhang, Runhui Li and Gong Zhang
BIG Cache Abstraction for Cache Networks.
The 37th IEEE International Conference on Distributed Computing Systems (ICDCS'17), Atlanta, Georgia, June 2017

Unit 335-337, Core Bldg – Hong Kong Science Park – Hong Kong
☎ *(+852) 9517 4226* • ✉ *lrhdiy@gmail.com* • *rhli.github.io*

2017 Runhui Li, Yuchong Hu and Patrick P. C. Lee
Enabling Efficient and Reliable Transition from Replication to Erasure Coding for Clustered File Systems.
Transaction on Parallel and Distributed Systems (TPDS) (to appear)

2015 Mingqiang Li, Runhui Li and Patrick P. C. Lee
Relieving Both Storage and Recovery Burdens in Big Data Cluster with R-STAIR Code.
2015 USENIX Annual Technical Conference (ATC'15) (Poster presentation), Santa Clara, CA, July 2015

2015 Runhui Li, Jian Lin and Patrick P. C. Lee
Enabling Concurrent Failure Recovery for Regenerating-Coding-Based Storage Systems: From Theory to Practice.
IEEE Transaction on Computers (TC), 64(7), pp. 1898-1911, July 2015

2015 Runhui Li, Yuchong Hu and Patrick P. C. Lee
Enabling Efficient and Reliable Transition from Replication to Erasure Coding for Clustered File Systems.
Proceedings of 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'15) (Regular paper), Rio de Janeiro, Brazil, June 2015
(AR: $50/229 = 21.8\%$)

2015 Runhui Li and Patrick P. C. Lee
Making MapReduce Scheduling Effective in Erasure-Coded Storage Clusters.
Proceedings of the 21st IEEE International Workshop on Local and Metropolitan Area Networks (LANMAN'15) (Invited paper), Beijing, China, April 2015

2014 Runhui Li, Patrick P. C. Lee and Yuchong Hu
Degraded-First Scheduling for MapReduce in Erasure-Coded Storage Clusters.
Proceedings of 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'14) (Regular paper), Atlanta, Georgia, June 2014
(AR: $56/181 = 30.9\%$)

2014 Silei Xu, Runhui Li, Patrick P. C. Lee, Yunfeng Zhu, Liping Xiang, Yinlong Xu and John C. S. Lui
Single Disk Failure Recovery for X-Code-Based Parallel Storage Systems.
IEEE Transaction on Computers (TC), 63(4), pp. 995-1007, April 2014

2013 Runhui Li, Jian Lin and Patrick P. C. Lee
CORE: Augmenting Regenerating-Coding-Based Recovery for Single and Concurrent Failures in Distributed Storage Systems.
Proceedings of 29th IEEE Conference on Massive Data Storage (MSST'13) (Short paper), Long Beach, CA, May 2013
(AR: $(14 + 15)/109 = 26.6\%$)

2011 Liping Xiang, Yinlong Xu, John Lui, Qian Chang, Yubiao Pan and Runhui Li
A Hybrid Approach to Failed Disk Recovery Using RAID-6 Codes: Algorithms and Performance Evaluation.
ACM Transaction on Storage, 7(3):11, Oct 2011

Unit 335-337, Core Bldg – Hong Kong Science Park – Hong Kong
☎ *(+852) 9517 4226* • ✉ *lrhdiy@gmail.com* • *rhli.github.io*

Honors and Awards

2015 The 45th IEEE/IFIP DSN Conference Student Travel Grant
2014 The 44th IEEE/IFIP DSN Conference Student Travel Grant
2011-2015 CUHK Postgraduate Studentship
2011 Undergraduate Excellent Thesis Award of USTC
2010 Citigroup Scholarship
2010 Excellent Thesis Award of Undergraduate Student Research Project of USTC
2009 Excellent Student Scholarship
2008 Excellent Student Scholarship
2007 Excellent Freshman Scholarship

Reference (Available upon Request)

Prof. Patrick P. C. Lee (Supervisor)
Department of Computer Science and Engineering,
The Chinese University of Hong Kong
E-mail: pclee@cse.cuhk.edu.hk

Prof. John C. S. Lui
Department of Computer Science and Engineering,
The Chinese University of Hong Kong
E-mail: cslui@cse.cuhk.edu.hk